

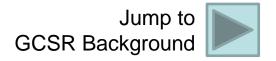
Geospatial Intelligence Standards Working Group

NSG Enterprise Vocabulary (NEV) Overview

Nathan J. Babcook, NGA,
GEOINT Content Semantic Resources
Program Lead
21 December 2017

NSG Core Vocabulary (NCV) Standard

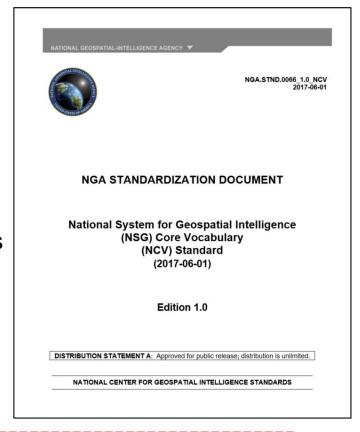
- Title: "National System for Geospatial Intelligence (NSG) Core Vocabulary (NCV) Standard"
- NCV Standard includes
 - NGA Standard document: NGA.STND.0066_1.0_NCV
 - NCV content: Technical artifacts (2 W3C-based SKOS encodings)
 - Resources available via REST API component of the NSGREG
- NSG Core Vocabulary Home Page: http://nsgreg.nga.mil/ncv
- Part of a family of GEOINT Content Semantic Resources
 - Abbreviation: "GCSR"
 - GCSR includes the NSG Application Schema (NAS), currently in use
 - Three new standards specify Semantic Web-enabled technical artifacts
 - NSG Core Vocabulary (NCV)
 - NSG Taxonomy (NTAX)
 - NSG Enterprise Ontology (NEO)





NCV Standard Covers NSG Vocabularies

- Outline of NCV Standard DRAFT
- Body of the Document
 - Section 1: Scope
 - Section 2: Conformance
 - Section 3: References
 - Section 4: Terms, Definitions, and Acronyms
 - Section 5: Vocabulary Specification
 - Section 6: Governance
- Annexes
 - Annex A (normative): Conformance
 - Annex B (normative): ICS Pro Forma
 - Annex C (normative): NSG Enterprise Vocabulary (NEV) Specification
 - Annex D (normative): Utility Ontologies for NCV
 - Annex E (*informative*): Inspecting the NCV Content
 - Annex F (*informative*): UML Primer





NCV Standard Scope: Terminology

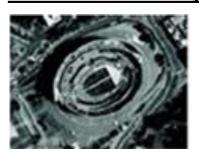
- Specification for a controlled vocabulary of defined terms intended to consistently describe GEOINT information shared in the NSG
 - A controlled vocabulary is a set of terms (CV elements) consisting of defined lexical items (for example: words, phrases, and/or abbreviations) from natural language that are collected and managed according to identified criteria. Examples: Glossaries, data dictionaries.
- Vocabulary content consists of:
 - Lexical items (names, phrases)
 - Natural-language definitions
- Leverages established terminology created by GEOINT Subject-Matter Experts (SMEs)
 - Selectively derived from the NSG Feature Data Dictionary (NFDD)
 - Selectively derived from content in the IR Registry



Term
"the Roman Coliseum"

Theory of Denotation: Words are related to the world through the mediation of concepts.

Real-world Object





Concept

Name: "Roman Coliseum" Alias: "Flavian Amphitheatre"

Entity type: Stadium

Definition: "An oval amphitheater in Rome built by Imperial order in the

first century C.E." Location: Rome, Italy Position: N41° 53'25",

E12º29'32"

Meaning: Terms represent concepts in the minds of language users.

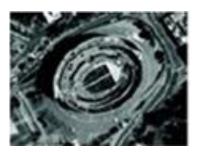
- Terms in different languages may represent the same concept.
- Controlled vocabularies ensure that terms are used and understood consistently by users.

Term

"the Roman Coliseum"

Theory of Denotation: Words are related to the world through the mediation of concepts.

Real-world Object





Meaning: Terms represent concepts in the minds of language users.

- Terms in different languages may represent the same concept.
- Controlled vocabularies ensure that terms are used and understood consistently by users.

Term

"the Roman Coliseum"

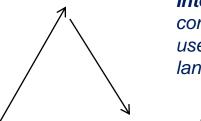
Theory of Denotation: Words are related to the world through the mediation of concepts.

Name: "Roman Coliseum"
Alias: "Flavian Amphitheatre"
Entity type: Stadium
Definition: "An oval amphitheater in

Rome built by Imperial order in the

first century C.E." Location: Rome, Italy Position: N41° 53'25",

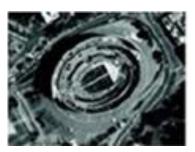
E12º29'32"



Concept

Interpretation: Concepts are constituents of mental models used by humans to apply language to the real world.

Real-world Object





signified (meaning)

Meaning: Terms represent concepts in the minds of language users.

- Terms in different languages may represent the same concept.
- Controlled vocabularies ensure that terms are used and understood consistently by users.

<u>Term</u>

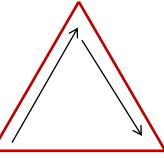
"the Roman Coliseum"

signifier (sign)

Theory of Denotation: Words are related to the world through the mediation of concepts.



Concept



Name: "Roman Coliseum"
Alias: "Flavian Amphitheatre"

Entity type: Stadium

Definition: "An oval amphitheater in

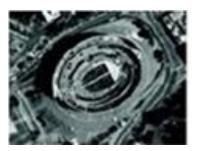
Rome built by Imperial order in the

first century C.E." Location: Rome, Italy Position: N41° 53'25",

E12º29'32"

Interpretation: Concepts are constituents of mental models used by humans to apply language to the real world.

Real-world Object



referent (object)



Concept

signified (meaning)

Meaning: Terms represent concepts in the minds of language users.

- Terms in different languages may represent the same concept.
- Controlled vocabularies ensure that terms are used and understood consistently by users.

Term

"the Roman Coliseum"

signifier (sign)

NCV Standard covers Vocabulary Terms

Theory of Denotation: Words are related to the world through the mediation of concepts.



Entity type: Stadium

Definition: "An oval amphitheater in Rome built by Imperial order in the

first century C.E."

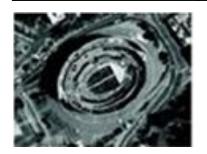
Location: Rome, Italy

Position: N41° 53'25". E12º29'32"

NAS & NEO Standards cover detailed **Concept models**

Interpretation: Concepts are constituents of mental models used by humans to apply language to the real world.

Real-world Object



referent (object)



Formal encoding of terms & concepts enables machines to help users share words and meanings.

Terms vs. Concepts

Terms

- Lexical items
 - Words, phrases, or abbreviations
- Defined in human language
- Enable understanding of a shared domain language
- Represented in flat lists
 - Dictionaries
- May be encoded
 - Language-specific strings
- Provided as a terminology resource
 - Simple(r)

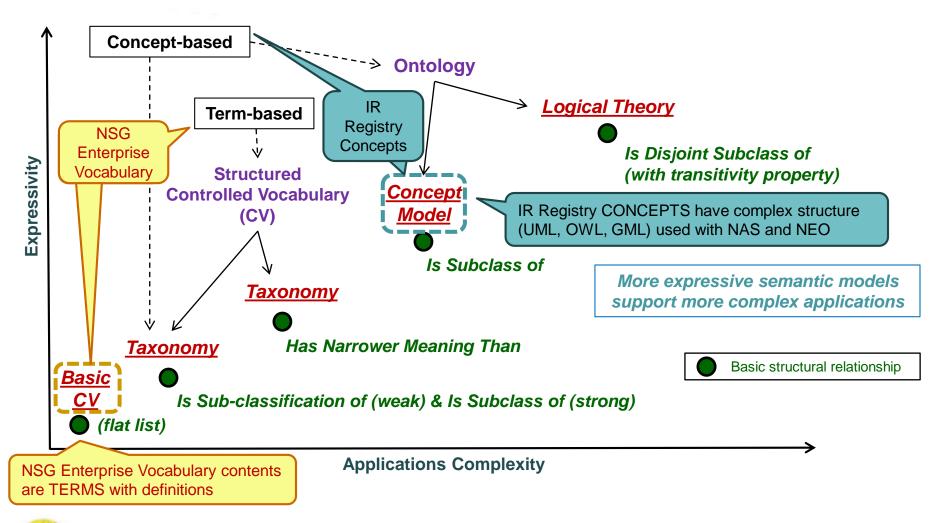
Concepts

- Structured items
 - Entity types and properties

 (abstractions of real-world objects)
- Defined in formal models
- Enable understanding of shared structured datasets
- Represented in schemas
 - UML
 - Ontologies (e.g., OWL)
- May be encoded
 - Logic-based languages
- Provided as a resource for data modeling and interoperability
 - Complex



Spectrum of Semantic Resources





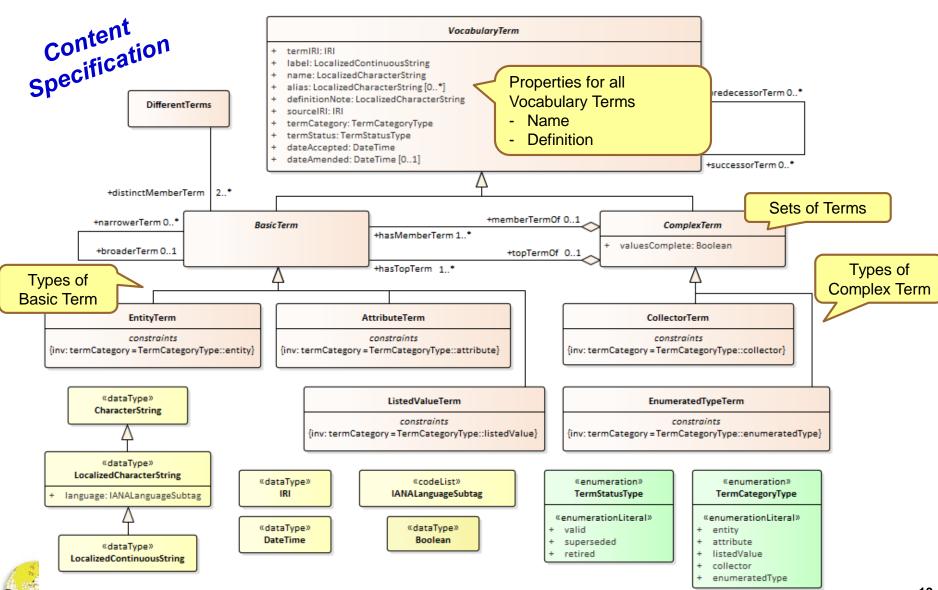
Vocabulary Specification

- NSG Vocabulary Information Model (Content Specification)
 - VocabularyTerm {Abstract} with subclasses BasicTerm & ComplexTerm
 - Represented in W3C Simple Knowledge Organization System (SKOS)
- IRIs* and Namespace (Content Identification)
 - Unique identifiers implemented with URI/URLs to locate resources
 - Related items collected within namespaces
- SKOS Representation (Content Encoding)
 - Simple Knowledge Organization System (SKOS) representation
 - SKOS technical artifacts are encoded in RDF/XML and N-Triples
 - Supports use of Vocabulary concepts in indexing, query, and search
 - Supports use of Vocabulary concepts with Linked Data

^{*}An Internationalized Resource Identifier (IRI) is a sequence of characters from the Universal Character Set (Unicode/ISO 10646) that forms an identifier for a resource. IRIs complement an older format, Uniform Resource Identifiers (URIs), which allows the use of only a subset of the ASCII character set to construct identifiers. Every URI is by definition an IRI. Reference: http://tools.ietf.org/html/rfc3987#section-3.1



NSG Vocabulary Information Model



Content NCV Namespace and Identifiers

Namespace

- A collection of conceptually related resources whose identifiers (IRIs) share a common prefix ("URI base").
- Each resource IRI belongs to a single namespace.
- Resources from different namespaces may be combined in the construction of a new resource ("re-use").

NSG Core Vocabulary Identifier

- A unique identifier (IRI) is assigned to the vocabulary as a whole:
 - http://api.nsgreg.nga.mil/vocabulary/ncv

Form of IRIs for NCV Components

- protocol "://" domain "/" resource-type "/" resource "/" concept
- Example: http://api.nsgreg.nga.mil/vocabulary/ncv/Building

Vocabulary Baselines and Vocabulary Term Status

- Authoritative NCV content is published with baseline metadata
- Non-versioned IRI retrieves the latest NCV content through the REST API
- Status for individual Vocabulary Terms is recorded for each term



Content Encoding

NSG Vocabulary Encodings

- Simple Knowledge Organization System (SKOS)
 - SKOS provides a Web-enabled representation for terminology
- The NSG vocabulary information model is formalized in SKOS.
 - NCV Basic Terms are represented by skos: Concept.
 - NCV Complex Terms are represented by skos: ConceptScheme.
 - SKOS properties are used for Term documentation
- NSG vocabularies have two encodings
 - RDF/XML: an XML-based format for encoding an RDF graph
 - N-Triples: a line-based, plain text format for encoding an RDF graph as a set of "Subject-Predicate-Object" (S-P-O) triples



NSG Enterprise Vocabulary (NEV)

- The NCV Standard, Annex C, presents the specification for the NSG Enterprise Vocabulary (NEV)
- NEV components use the same NSG vocabulary information model as NCV content
- Content of NEV components is related to selected contents of the IR Registry
- NEV = NCV + NEV components

NSG Core Vocabulary (NCV) Standard, Edition 1.0

Annex C – NSG Enterprise Vocabulary (NEV) Specification (Normative)

C.1 Introduction

The NSG Enterprise Vocabulary (NEV) is a collection of controlled vocabularies including the NSG Core Vocabulary (NCV), which conform to the NSG vocabulary information model and which may be used together to describe geospatial intelligence (GEOINT) shared in the National System for Geospatial Intelligence (NSG).

This Annex specifies the IRIs for NEV resources that are based on the content of the Information Resources Registry of the NSG Standards Registry (Section C.2). Content will be published in the future.

This Annex also describes the requirements (Sections C.3, C.4, C.5, C.6) for any controlled vocabulary to be a component of the NEV.

C.2 IRIs

IRIs for the NEV and its components are constructed in accordance with the following pattern specified in the NCV Standard, Section 5.5.2, with each vocabulary having its own resource abbreviation, based on the name of the resource:

protocol ":/f" domain "/" resource-type "/" resource "/" concept

In this pattern, each component is case-sensitive and determined as follows:

- protocol always 'http'
- domain always 'api.nsgreg.nga.mil'
- resource-type always 'vocabulary'
- resource abbreviation for a specific resource (e.g., 'ncy' for the NCV; 'ney' for the NEV)
- concept designates an individual vocabulary term (e.g., 'Aerial', 'Building', 'MountainPass')

The individual components are concatenated into a single string as specified by the pattern (above), to form the IRI that designates the associated vocabulary component.

The IRI for the NEV is http://api.nsgreg.nga.mil/vocabulary/nev.

The identifiers established for component vocabularies of the NSG Enterprise Vocabulary are listed in Table 19.

Table 19 – IRIs for NEV and Component Vocabularies

Name	Resource IRI		
Codelist	http://api.nsgreg.nga.mil/vocabulary/codelist		
Belief System	http://api.nsgreg.nga.mil/vocabulary/belief-system		
Linguistic Entity	http://api.nsgreg.nga.mil/vocabulary/linguistic-entity		
Physical Quantity	http://api.nsgreg.nga.mil/vocabulary/physical-quantity		

NSG Enterprise Vocabulary (NEV) IRI:

http://api.nsgreg.nga.mil/vocabulary/nev



NEV Components for IR Resources

- Information Resources (IR) Registry resources specify many specialized NSG domain concepts
- NEV component vocabularies define terminology for IR Registry content
 - One vocabulary for each IR Register
- Vocabulary terms don't capture the logical complexities of the IR Registry concepts
 - Terms are lexical items with definitions
 - Terms are entry points for human understanding
 - Encoded terms may be used for indexing, tagging, and term search
- IR Registry concepts are based on complex information models
 - Conceptual models are used to represent structured knowledge
 - Represented in UML, GML, OWL



Example: QMV Term for a QM Concept

Quality
Measures
Vocabulary
term (lexical
information)

name: Rate of Incorrect Attribute Values

termIRI: http://api.nsgreg.nga.mil/vocabulary/quality-measure/RateIncorrectAttributeValues

label: RateIncorrectAttributeValues

definitionNote: Definition: The number of attribute values where incorrect values are

assigned in relation to the total count of attribute values. Value Type: Real (Singleton)

termCategory: attribute

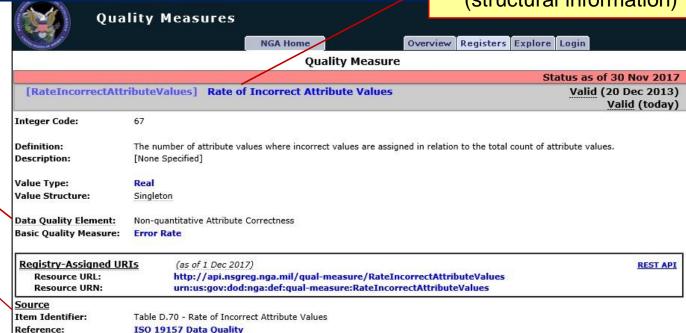
sourcelRI: http://nsgreg.nga.mil/ir/view?i=300099

20 December 2013

Context Added

termStatus: Valid

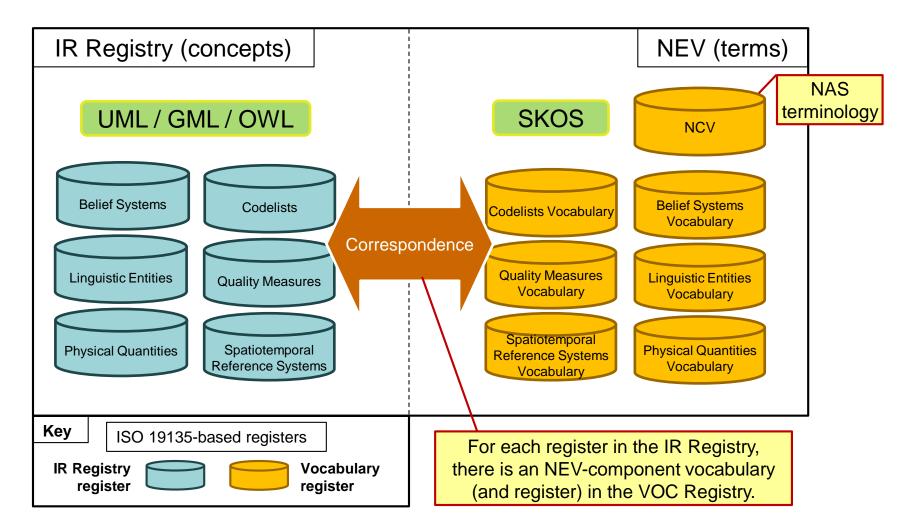
Reference Date: Similarity: Correlated IR QM concept (structural information)



Additional structural information



IR Registers Paired with NEV Components





NEV Vocabularies vs. IR Registry Concepts

Vocabulary Name	Scope	IR Registry Item Classes
Belief Systems Vocabulary	A vocabulary for perspectives about the world held by human beings (past or present), often as part of a shared subculture.	Belief System
Codelists Vocabulary	A vocabulary for sets of value domains that are used to specify geospatial data and metadata.	Datatype & Listed Value
Linguistic Entities Vocabulary	A vocabulary for systems of sounds, characters, symbols, or signs used to express or communicate meaning or messages between humans.	Linguistic Entity
Physical Quantities Vocabulary	A vocabulary for properties having a magnitude that can be expressed by means of a number (physical value) and a reference quantity (referred to as a "unit of measure").	Physical Quantity, Unit of Measure, Unit Multiple
Quality Measures Vocabulary	A vocabulary for evaluation procedures used to determine the degree to which the inherent characteristics of data fulfil requirements.	Quality Measure, Basic Quality Measure, Quality Measure Parameter
Spatiotemporal Reference Systems Vocabulary	A vocabulary for reference systems and related components that are used to specify spatial and/or temporal aspects of geospatial data and metadata.	SRS Coordinate Reference System, SRS Coordinate System, SRS Coordinate System Axis, SRS Datum, SRS Prime Meridian, SRS Ellipsoid, SRS Coordinate Operation, SRS Operation Method, SRS Operation Parameter, Spatiotemporal Direction



IRIs for NEV Components

Vocabulary Name	Code	Resource IRI
Codelists Vocabulary	CDV	http://api.nsgreg.nga.mil/vocabulary/codelist
Belief Systems Vocabulary	BSV	http://api.nsgreg.nga.mil/vocabulary/belief-system
Linguistic Entities Vocabulary	LEV	http://api.nsgreg.nga.mil/vocabulary/linguistic-entity
Physical Quantities Vocabulary	PQV	http://api.nsgreg.nga.mil/vocabulary/physical-quantity
Quality Measures Vocabulary	QMV	http://api.nsgreg.nga.mil/vocabulary/quality-measure
Spatiotemporal Reference Systems Vocabulary	SRV	http://api.nsgreg.nga.mil/vocabulary/coord-ref-system



Publication

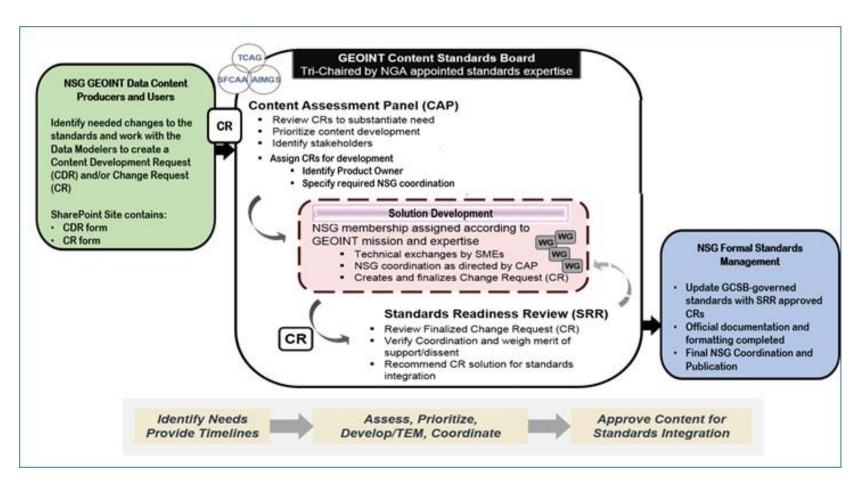
- GCSR Standards (NAS, NEO, NTAX, & NCV) and associated technical artifacts are published in the <u>NSG-unique Standards</u> <u>Register</u> of the NSG Standards Registry
- NSG Core Vocabulary (NCV) Standard
 - "National System for Geospatial Intelligence (NSG) Core Vocabulary (NCV) Standard" (NGA.STND.0066_1.0_NCV)
 - MS Word document containing the vocabulary specification and the conformance test suite
 - URL: http://nsgreg.nga.mil/doc/view?i=2617

Managed content

- NCV Register (online)
- Technical artifacts
 - RDF/XML encoding of the NSG Core Vocabulary (content)
 - N-Triples encoding of the NSG Core Vocabulary (content)
- Online resources accessible through the REST API component of the NSG Standards Registry
 - Resources for individual terms, as well as for the vocabulary as a whole
 - HTTP content negotiation used to specify media type (RDF/XML or N-Triples)



Governance



- The GEOINT Content Standards Board (GCSB) manages the GCSR Standards
- The GCSB Operations Guide specifies GCSB management roles and processes.
- The GCSB oversees publication of the GCSR Standards and technical artifacts.

Point of Contact

Nathan J. Babcook (Nathan.J.Babcook@nga.mil)

- GEOINT Content Semantic Resources Program Lead,
 NGA CIO & IT Services Directorate
- Chair, GWG Metadata Focus Group (<u>mfgchair@nga.mil</u>) and Application Schemas for Feature Encoding (ASFE) Focus Group (<u>asfechair@nga.mil</u>)



BACKGROUND: GEOINT CONTENT SEMANTIC RESOURCES (GCSR)



Types of Semantic Resources

Semantic Resource Type	Definition
Basic Controlled Vocabulary (CV)*	A set of terms (CV Elements) consisting of defined lexical items (for example: words, phrases, and/or abbreviations) from natural language that are collected and managed according to identified criteria. Examples: Glossaries, data dictionaries.
Taxonomy	A hierarchical (tree) structure consisting of either (1) a collection of terms (term-taxonomy) or (2) a collection of concepts, arranged into a subordination tree. Strict ("strong") and loose ("weak") taxonomies are distinguished based on the type of hierarchical relationship used (subclass <i>vs.</i> sub-classification , respectively). Examples: Library of Congress classification (weak taxonomy); Linnaean classification of animals (strong taxonomy).
Thesaurus	A collection of defined terms structured by several standardized lexical relationships, including: narrower-than and broader-than (for loose/weak hierarchies); used-for (to indicate synonyms); related-to (for relationships); and preferred-term may be indicated. Examples: Roget's Thesaurus, Thesaurus.com.
Concept Model	A complex formalization of concepts for representing domain objects using classes, attributes, hierarchical (<i>subclass-of</i>) relationships, and other relationships (associations) between classes of objects. The structure of a concept model is a graph rather than a tree. Example: UML models.
Ontology (Logical Theory)	A logical theory for a domain (part or all of the world), including formally defined concepts and properties, with restrictions and optionally rules, in which the structural relationships (including <i>subclass-of</i> , <i>equivalence</i> , and <i>disjoint-with</i>) are defined in a formal logic (either axiomatically or in a rule-based formulation). "An ontology is an explicit specification of a conceptualization." (Thos. R. Gruber) Examples: Basic Formal Ontology (BFO), Suggested Upper Merged Ontology (SUMO); Friend of a Friend (FOAF).
	* "Controlled Vocabulary" may be used generically to include dictionaries, term-taxonomies, and thesauri. "Basic CV" is used to distinguish simple glossary CVs from more complex CVs



that include relationships between terms.

New GCSR Standards & Web-enabled Resources

NSG Taxonomy (NTAX)

- Entity Class hierarchy for classification of GEOINT shared on the NSG
- Semantic Web-enabled using Web Ontology Language (OWL 2)
- Conceptual content based on NSG Application Schema* entity types

NSG Core Vocabulary (NCV)

- Controlled terminology for describing GEOINT shared in the NSG
- Web-enabled using Simple Knowledge Organization System (SKOS)
- Specification for the NSG Enterprise Vocabulary (NEV)**

NSG Enterprise Ontology (NEO)

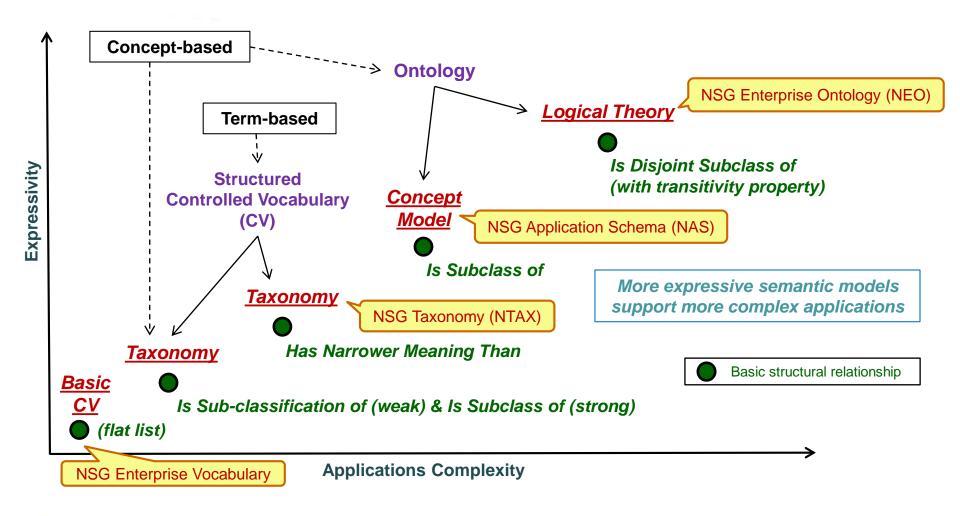
- Large multi-domain GEOINT ontology with a complex design-space
- Web-enabled using Web Ontology Language (OWL 2) + SKOS
- Content derived from the NAS using ISO rule-based approach

NSG Enterprise Thesaurus (NET)

- Potential to link other SKOS-encoded terminologies across the NSG
- NCV can be related to other terminologies (as synonyms; broader/narrower terms; related terms)
 - * For NAS, see http://nsgreg.nga.mil/nas/
 - ** In NCV Standard, Ed. 1.0, Annex C



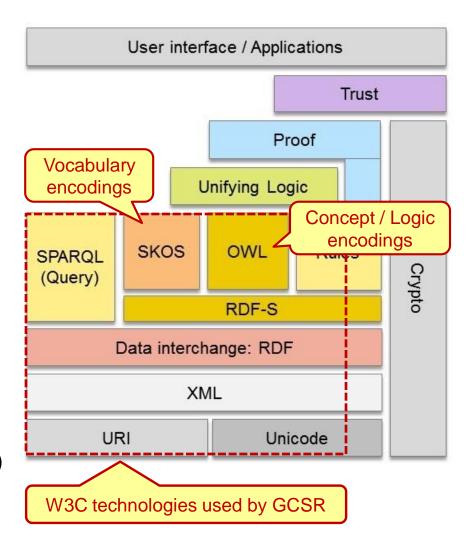
Spectrum of Semantic Resources





Technologies in the Semantic Web "Stack"

- "Layered" use of multiple standards to support interoperability of data with machine-processible semantics on the internet
- Foundation: URIs to identify Web resources
- Structure: XML for syntax
- Semantics: World Wide Web Consortium (W3C) standards ("Recommendations")
 - Resource Description
 Framework (RDF)
 - RDF Schema (RDF-S)
 - Web Ontology Language (OWL)
 - Simple Knowledge Organization System (SKOS)





Uses of Semantic Resources in Applications

Applications	CV	Thesaurus	Taxonomy	Concept Model	Ontology
Terminology services (e.g., definitions, synonyms)	X	X			
Labeling (e.g., menus, interfaces)	X	X			
Categorization (e.g., of documents or data instances)			X	X	X
Indexing (e.g., multimedia collection resources)	Χ	X	X	X	Χ
Search (Keyword)	X	X	X	X	X
Search (Semantic)		X	X	X	Χ
Navigation		X	Χ	X	Χ
Content representation (e.g., domain ontology)			Χ	X	Χ
Constraint assertions and checking			Χ	X	Χ
Web Services			X	X	Χ
Mapping models (e.g., to a reference ontology)				X	Χ
Data integration (instance-level)				X	Χ
Unified querying over heterogeneous data					Χ
Inference (derive new information about data)					Χ
Semantics for Linked Data (M2H; M2M)					Χ



RETURN TO NEV OVERVIEW

